Clinical Microbiology and Infection 22 (2016) 600–606

Contents lists available at ScienceDirect

Clinical Microbiology and Infection

journal homepage: www.clinicalmicrobiologyandinfection.com

Review

Use of systems biology to decipher host–pathogen interaction networks and predict biomarkers

A. Dix [1], S. Vlaic [1, 2], R. Guthke [1], J. Linde [1, *]

[1)] *Systems Biology/Bioinformatics, Leibniz Institute for Natural Product Research and Infection Biology, Hans-Knöll-Institute*
[2)] *Department of Bioinformatics, Friedrich-Schiller-University, Jena, Germany*

ARTICLE INFO

Article history:
Available online 22 April 2016

Keywords:
Biomarker identification
Disease modules
Drug target prediction
Host–pathogen interaction
Network modelling
Personalized medicine
Systems medicine

ABSTRACT

In systems biology, researchers aim to understand complex biological systems as a whole, which is often achieved by mathematical modelling and the analyses of high-throughput data. In this review, we give an overview of medical applications of systems biology approaches with special focus on host–pathogen interactions. After introducing general ideas of systems biology, we focus on (1) the detection of putative biomarkers for improved diagnosis and support of therapeutic decisions, (2) network modelling for the identification of regulatory interactions between cellular molecules to reveal putative drug targets and (3) module discovery for the detection of phenotype-specific modules in molecular interaction networks. Biomarker detection applies supervised machine learning methods utilizing high-throughput data (e.g. single nucleotide polymorphism (SNP) detection, RNA-seq, proteomics) and clinical data. We demonstrate structural analysis of molecular networks, especially by identification of disease modules as a novel strategy, and discuss possible applications to host–pathogen interactions. Pioneering work was done to predict molecular host–pathogen interactions networks based on dual RNA-seq data. However, currently this network modelling is restricted to a small number of genes. With increasing number and quality of databases and data repositories, the prediction of large-scale networks will also be feasible that can used for multidimensional diagnosis and decision support for prevention and therapy of diseases. Finally, we outline further perspective issues such as support of personalized medicine with high-throughput data and generation of multiscale host–pathogen interaction models. **A. Dix, CMI 2016;22:600**

Introduction

Systems biology aims at understanding and modelling complex biological systems as a whole. Systems biology approaches may focus on single or multiple levels, including genes, proteins, cells, tissues, whole organisms or populations. This area of research has emerged at the beginning of the 21st century [1–3]. It is characterized by a strong connection of wet lab experiments and computational analysis, where the analysis and modelling of experimental data results in new hypotheses which lead to new experiments [1].

Typically systems biology approaches are categorized into bottom up and top down, where the bottom describes the interaction of molecules and the top is the holistic view on the system. Bottom-up approaches aim for elucidating the interactions of cell components by submodel aggregation [4]. The view from the top is created by genome-wide analysis and thus by the so-called omics technologies (genomics, transcriptomics, proteomics, metabolomics) [5].

Microarrays and RNA sequencing (RNA-seq) are the two techniques mainly used for transcriptome wide gene expression analysis. In contrast to microarrays, RNA-seq has a higher sensitivity in measuring genes with low abundances and allows the identification of novel transcripts as well as sequence variations [6]. Moreover, RNA-seq facilitates expression analysis for non-model organisms, including pathogenic microorganisms, since the expensive step of manufacturing species-specific arrays is not necessary. The analysis of proteomes at high sensitivity is facilitated by advances in mass spectrometry–based methods [7], which superseded 2-D gel-based proteomics [8]. These methods enable researchers to determine the quantity of proteins and metabolites.

* Corresponding author. J. Linde, Beutenbergstraße 11a, 07745 Jena, Germany
E-mail address: joerg.linde@leibniz-hki.de (J. Linde).

http://dx.doi.org/10.1016/j.cmi.2016.04.014

Systems Biology of Host–Pathogen Interactions

From a clinical perspective, systems biology of host cells, pathogen cells and their interaction with each other are of special interest. In both players, these interactions result in the triggering of signalling cascades, which may change the activity of proteins and thus directly result in a phenotypic response. These changes happen on different molecular levels and are quantified using omics and image data, which computational analysis and modelling are challenging [9–11]. Often gene expression data are used, as it gives a representative picture of the response of an organism to an environmental change. In fact, gene expression data of either the host or the pathogen during infection related conditions or *in vivo* have strongly contributed to our knowledge about virulence factors, biomarkers, host immunity and dynamics of infection [12–14]. Going beyond the analysis of single-species transcriptomes, dual trancriptomics simultaneously measure the gene expression of the host and the pathogen during an infection. Here, processing and measurement of transcript abundances of both host and pathogen are performed simultaneously, while species-specific expression is determined *in silico* [15,16]. There are a few examples where dual transcriptomics were performed using dual microarray analysis [17,18]. In addition, the advent of species-independent RNA-seq platforms allows for high-quality dual RNA-seq [15,19].

Besides analyses restricted to the transcriptome, systems-level approaches have been used to elucidate the virulence of microorganisms like *Salmonella* [20–22], *Yersinia pestis* [23] or *Mycobacterium tuberculosis* [24]. These procedures also take metabolic pathways and gene–protein interactions into account. Multiple databases have been set up to support the analysis of host–pathogen interactions [9]. While some are specific to types of organisms (e.g. bacteria or fungi) [25,26], others cover broader ranges of pathogens and are only limited by the available data [27–29].

Based on genome-scale metabolic modelling, analyses of the gut microbiome have been performed, revealing interaction of the microbial ecosystems, their interactions with the host and associations with disorders [30–34], also addressing symbiotic relationships [35]. Recently the tool CASINO (Community And Systems-level INteractive Optimization) has been developed, which allows examination of microbe–microbe, microbe–host and diet–microbe interactions based on metabolic modelling [36].

In systems biology, there are three widely accepted approaches for the elucidation of key molecules and their interactions within and between organisms (Fig. 1). First is the identification of key molecules, i.e. biomarkers, to monitor or predict disease progression and the condition of the host and thus ultimately support the therapeutic decision making. Second is the network inference for the identification of relations between molecules, thereby uncovering potential drug targets. Third is the identification of disease modules, i.e. groups of molecules and interactions associated with a certain phenotype. In the following, we describe these approaches in detail and provide exemplary findings in these research areas.

Biomarker Discovery

There are several definitions of biomarkers, which are discussed by Strimbu and Tavel [37]. Some restrict biomarkers to application areas like diseases, pathogenecity or pharmacology, while other definitions are more general. Briefly, a biomarker can be considered as an objectively measurable medical sign that indicates a condition or state. Biomarkers can be of any kind, including omics data, like DNA/RNA sequences, proteins, mutations (SNPs) (Fig. 1) and other attributes of an organism, e.g. an increased heart rate or body temperature. With regard to host–pathogen interactions, biomarkers can originate from both players. For example, the detection of the *Aspergillus* cell wall component galactomannan in the

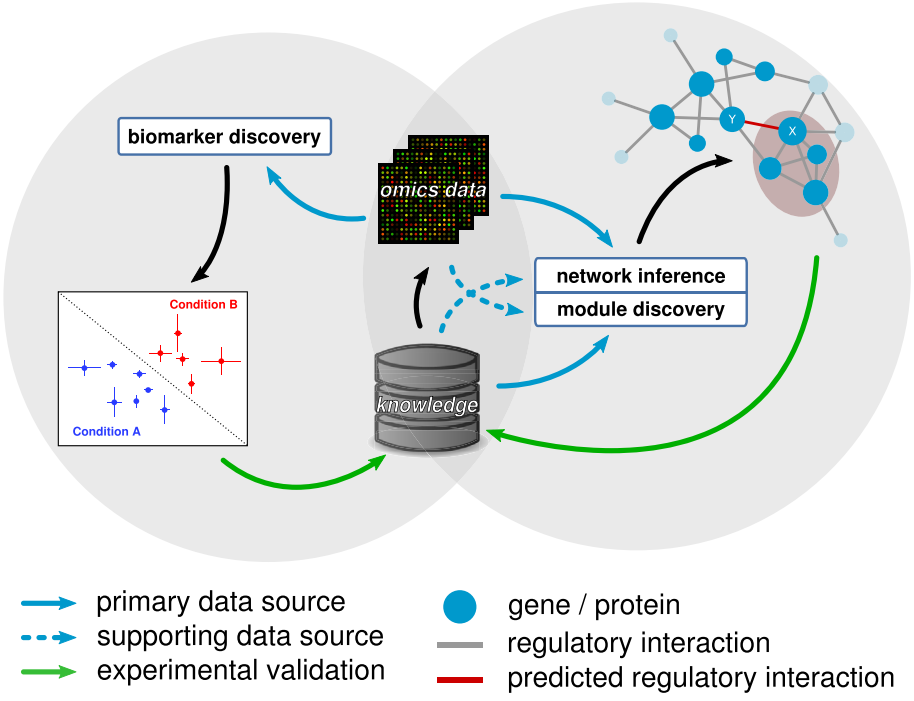

Fig. 1. Overview of the relations of systems biology approaches for biomarker discovery, network inference and module discovery. With the help of omics data, biomarkers can be identified that indicate patient conditions and may serve as diagnostic or prognostic marker. Omics data are used for network inference, where, with support of prior knowledge, regulatory networks are reconstructed. These networks predict novel molecular interactions as well as important key regulatory molecules, which may become novel therapeutic targets. The interaction network derived from prior knowledge forms the basis for the discovery of disease modules. Creation of hypotheses in this field is supported by omics data and/or known associations between diseases and molecules. All three approaches require experimental validation of the generated hypotheses, which then serve as new prior knowledge for future experiments and analyses.

serum of the host has emerged as an important biomarker for supporting the diagnosis of invasive aspergillosis [38]. An example for a host-specific marker is an increased procalcitonin level, which is used as an indicator for bacterial infections, including sepsis [39]. Furthermore, procalcitonin correlates with the severity of the infection and thus also serves as a prognostic biomarker for severe sepsis and septic shock [40].

With respect to omics data, the techniques for the detection of biomarkers are numerous. A common approach is the analysis of differential regulation based on statistical significance [41,42]. The required tools for this task are fast and well established. However, this rather simple approach has some drawbacks. Especially for large-scale quantifications (omics), the resulting set of potential biomarkers may contain large redundancies and thus also many false-positive findings. Therefore, subsequent validation experiments may be expensive, as more candidates have to be verified. Moreover, this method does not consider combinatorial effects. A more complex but promising approach is the incorporation of machine learning techniques [43–46]. These techniques comprise classification and/or regression algorithms, including, but not limited to, random forest [47], support vector machines [48], k-nearest neighbours [49] and linear discriminant analysis [50]. While the biomarker identification using machine learning methods requires a higher computational effort, they often reduce the set of potential biomarkers to the most relevant candidates. Additionally, some of them, e.g. random forest, take interactions of variables into account, which means that combinatorial effects of biomarker candidates are considered [51].

The identification of transcriptional biomarkers is of particularly great interest because the regulation of gene expression reflects the cellular response and the methods of measurement are well established, fast and cost-effective. For example, gene signatures that distinguish between latent and active tuberculosis [52] as well as regulatory patterns to distinguish bacterial and fungal whole-blood infections have been identified [44].

Apart from protein coding transcripts, another focus of the endeavours of biomarker discovery lies with noncoding RNAs (ncRNAs), such as microRNAs (miRNAs) or Piwi-interacting RNAs (piRNAs). Up to now, miRNAs are the best-investigated ncRNAs [53] and have been successfully linked to infectious diseases such as active tuberculosis [54], hepatitis B virus infections [55] and the discrimination of the systemic inflammatory response syndrome with sepsis [56].

Although transcriptional biomarkers can improve the diagnosis and prognosis of infectious diseases, proteins represent the functional level of gene expression, and thus they may better reflect cellular responses. Therefore, they are the primary target point for biomarker studies. The identification of proteomic biomarkers, however, is more complex. Proteins have a larger diversity as a result of posttranslational modifications. Furthermore, they can form multiprotein complexes, which may alternate the function of the individual proteins. Nevertheless, there are successful advances in infection biology. In tuberculosis research, protein biomarkers were recently found that may allow tuberculosis diagnosis even in immunocompromised patients [57].

Another area of biomarker analysis is the field of genome-wide association studies for SNP identification. These studies aim to find connections between SNPs and disease phenotypes. Many SNPs in protein coding regions are linked to severe health risks. SNPs also have been associated with the susceptibility to infections and diseases, like pneumonia in H1N1 influenza infection [58] or sepsis [59]. The connections between SNPs and diseases can be drawn using DNA microarrays, which deliver the variant information of ≈ 1 million SNPs or more [60]. Due to the emergence of next-generation sequencing, in particular DNA-seq, novel SNPs can be detected *de novo*, i.e. without a reference genome, allowing SNP detection at high precision [61].

Molecular Network Modelling

In systems biology, the ultimate aim is to understand the molecular mechanisms underlying all biological processes on a systems level by means of mathematical modelling. These mechanisms are commonly represented in a network composed of nodes denoting molecules (e.g. proteins, DNA, RNA or metabolites) and edges representing an interaction between the connected nodes (e.g. protein–protein, protein–DNA). One can distinguish two broad areas in network science: first, the generation and structural analysis of interaction networks based on omics data supported by prior knowledge (Fig. 1) and second, the integrated analysis of interaction networks regarding the discovery of disease associated modules supported by prior knowledge and omics data (Fig. 1).

Automated inference of gene regulatory networks

With respect to mathematical modelling, a widely accepted approach for the generation of interaction networks is the automated reverse engineering. Here, perturbations are used to change the abundance of intracellular molecules, which is captured by omics data. The model is then fitted to the data, minimizing the deviation between the predicted and the observed molecule abundances. In the case of gene expression microarray or RNA-seq data, these interaction networks are also termed gene-regulatory networks (GRNs). Although the availability of omics data is increased by decreasing costs, advancing standard operating procedures and reliability of the measurements in conjunction with growing public data repositories, there are still central problems to the automated inference of GRNs. Usually the number of available measurements is much lower compared to the number of genes in the mathematical model underlying the GRN. This results in an underdetermined system with many equally good models. In general, there is a relation between the complexity of the model in terms of the number of free parameters per gene, the data required to explain the observed systems behaviour and the quality of the inferred GRN [62].

On the basis of this trade-off, a variety of approaches to the inference of GRNs have been proposed. They can be roughly grouped into two classes according to the number of genes integrated in the final network. On the one hand, there are approaches for the inference of large-scale or even genome-wide networks. These algorithms are computationally fast, using simplified, usually static mathematical models that display either correlation between genes predicting undirected edges (e.g. correlation-based algorithms [63] and information theory-based algorithms [64–67]) or causal relations among genes using weighted, directed edges (e.g. regression-based algorithms [68–70]). The resulting large-scale networks are usually used to study the topology of the network graph. Quantifiable measures such as (1) node-degree distribution [71], (2) betweenness centrality and (3) the clustering coefficient [72] give rise to (1) signal distributing hub genes, (2) central genes with a large influence on the transfer of signals through the network and (3) the underlying modular structure, respectively. For example, network inference based on expression data of the human pathogen *Candida albicans* was successfully applied for the identification of hub genes as new potential drug targets [73].

On the other hand, there are approaches for the inference of small-scale networks. These algorithms use more sophisticated, often dynamic mathematical models introducing biologically relevant aspects, such as nonlinearity, in the relation between

genes or the activation of genes, e.g. by detailed kinetic knowledge [74]. Common algorithms use models based on Boolean logic [75], Bayesian networks [76], difference equation systems [77] and linear [78] or nonlinear differential equation systems [79]. In the human pathogenic fungus *Aspergillus fumigatus,* the inference of a dynamic GRN was used to predict novel interactions between transcription factors and target genes relevant for the regulation of iron homeostasis [80].

Network inference has been mostly applied to predict GRNs of either the host or the pathogen. However, the advent of dual transcriptomics provides the possibility to infer molecular host–pathogen interaction networks, giving rise to the molecular mechanisms affected upon host–pathogen interaction [16]. In a pioneering study, inference of the network underlying *Candida albicans* infection of murine dendritic cells revealed two molecular host–pathogen interactions indicating an interaction between the inflammatory receptor PTX3 and the fungal transcription factor HAP3, which were experimentally validated [19].

Network analysis uncovering disease modules

Recent advances in the field of network science suggest the existence of disease modules in molecular networks, i.e. a group of proteins residing in the same neighbourhood of the underlying interaction network [81]. Several studies that identify modules underlying diseases such as obesity and type 2 diabetes [82], asthma [83] and inflammatory and malignant diseases [84] demonstrated that the identified disease modules partially overlap, sharing molecular mechanisms as well as proteins. It was further shown that the degree of overlap correlates with biological similarity, disease symptoms and increased evidence of comorbidity [85]. The analysis of disease modules, in conjunction with diverse data types (e.g. genomics, genome-wide association studies, gene expression, gene–disease association, clinically relevant information), has already proven a successful strategy to the discovery of affected molecular mechanisms, biomarkers and new drug targets as well as the repositioning of drugs. For example, analysis of the disease module underlying asthma revealed GAB1 signalling pathway as a novel modulator of asthma and successfully linked the pathway to glucocorticoids as a therapeutic agent [81]. In a study of breast cancer, the combined use of an interaction network, disease-specific gene expression data and patient survival data revealed a module enriched in Aurora B signalling and kinetochore-associated genes, whose expression was successfully validated to anticorrelate with patient survival [86]. Translation of the concept of disease modules on the host–pathogen relationship still remains problematic due to a lack of genome-wide interaction networks comprising both host and pathogen. Currently such models are restricted to gene subsets [87] or describe gene islands [88,89].

Outlook

Systems medicine

The analysis of host–pathogen interactions culminate in medical applications to improve diagnostics and therapies, thereby increasing survival chances and reducing adverse effects. For this aim, the results of systems biology studies have to be evaluated in clinical trials. The focus on medical applications based on discoveries and methods of systems biology, including the creation of associations between gene/protein behaviour to diseases and disorders, is the topic of systems medicine [90–92]. This interdisciplinary approach aims to extend the spectrum of available therapeutic and diagnostic molecules by systems-wide analysis.

The analysis can be further enhanced by incorporating clinical features, as shown by Reyes-Palomares *et al.* [93]. They combined disease network models with pathophenotype data and thereby found new connections between genes and diseases. This approach can improve the understanding of the pathologic process and the molecular mechanisms behind the host response. Moreover, a comprehensive analysis of already known biomarkers to identify associations to conditions different from the initial studies may contribute to a broader knowledge and an extended range of applications. Increased levels of glycoprotein acetylation have been determined as biomarker for incidence and mortality of cardiovascular disease [94,95]. In addition to this, Ritchie *et al.* [96] discovered that these elevated levels serve as marker for increased long-term risk of infection, especially septicaemia and pneumonia. In conclusion, the identification of biomarkers to develop patient- and tissue-specific therapeutic solutions is one of the key aspects in systems medicine [97]. Similarly, the elucidation of interactions of genes and/or proteins by network analysis is crucial to reveal disease processing [98].

Personalized medicine

The aim of personalized medicine is to improve individual treatment and therapy in order to maximize benefit and minimize harm of patients [99]. The focus here is to account for patient-individual characteristics, including molecular markers (e.g. genetic markers). Thus, personalized medicine is strongly connected to systems medicine research as the biomarkers are needed to monitor the influence of a treatment to an individual. In fact, genomics, transcriptomics, proteomics and metabolomics approaches are of indispensable need to identify such biomarkers using unbiased approaches [100].

As a result of rapidly decreasing costs of omics technologies, especially in the field of next-generation sequencing, there is no long way to gain personal omics data of individuals at different molecular levels. In fact, the dream of sequencing a human genome for US$1000 is close to coming true [101]. Because most drug targets are proteins, the identification and utilization of proteomic markers is most common in clinical settings [102]. Therefore, personalized medicine heavily relies on quantitative proteomics. Frequently biomarker prediction is initiated by a genome-wide transcriptomic screening based on microarray or RNA-seq. This screening is followed by validation of selected genes using quantitative real-time PCR and quantification of the proteins coded by these genes (e.g. by Western blot test or ELISA). In some cases, the validation of transcriptome-derived biomarkers on protein level fails. Possible reasons are posttranscriptional and translational regulations, protein modifications and different half-lives of proteins and RNA. Comparing protein and RNA levels, a squared Pearson correlation coefficient of ≈ 0.4 has been reported, showing a moderate relation between the expression levels [103,104]. Therefore, the generation of proteomic data at large scale is of particular importance for clinical studies.

The individual omics data allow two ways of application to personalized medicine. Static personal and clinical information (e.g. previous diseases) in combination with static biomarkers (e.g. genetic markers) [105,106] will help to stratify the personal risk of an individual gaining infection, suffering from a disease or responding to a therapy. Furthermore, dynamic clinical data in combination with dynamically changing biomarkers detected from omics data will help support therapeutic decisions such as the end of a therapy or when to change drug treatment. Moreover, such dynamic biomarkers can predict the outcome of an infection. One major research topic in system medicine will be the combination of biomarkers from multiple omics data together with

clinical biomarkers to ultimately improve predictions and diagnosis [100,107].

Translational system biology

Dynamic biomarkers are a starting point for 'translational systems biology.' This phrase was coined by Vodovotz and colleagues [108,109] and describes the utilization of computer simulation to optimize the design of drugs, as well as to simulate clinical trials and the results of a therapy on individuals. Thus, translational systems biology applies modelling and engineering approaches with the aim of optimizing clinical practice [110]. Especially, the extensive application of detailed mathematical modelling makes translational systems biology a promising research area since they not only predict a certain result but also contain information about why this result is happening [111]. Thus, these models will give clinicians a starting point for the change of a therapy. One particularly important direction of research is to set up models which are not only built upon data from one (molecular) level but models including data from different levels (i.e. molecular, tissue, organ, whole body) [10]. Examples of success stories including multiscale models are models which allow modulation of acute inflammation [109], computer-optimized antibiotic treatment for tuberculosis [112] and a novel mechanism for insulin resistance [113]. Also, in this field, the future task is to incorporate not only data from different molecular or macromolecular levels in the model but also personal data which may affect the system's response to stimuli *via* differences in the initial conditions.

In summary, systems biology paved the way for the concept of systems medicine (or 'P4 medicine'—predictive, preventive, personalized and participatory) towards 'precision medicine' [114]. The integration of omics, clinical data and analysis of large patient data sets has the potential to provide the basis for better molecular understanding of diseases and as a consequence to design novel and improved methods for diagnosis, prevention and therapy of diseases [115].

Transparency Declaration

This work was supported by the Deutsche Forschungsgemeinschaft CRC/Transregio 124 'Pathogenic fungi and their human host: Networks of interaction' subprojects B3 (to AD and RG) and INF (to JL) and the German Federal Ministry of Education and Research (BMBF, 'InfectControl,' subproject FINAR) (to SV). All authors report no conflicts of interest relevant to this article.

References

[1] Kitano H. Systems biology: a brief overview. Science 2002;295:1662–4.
[2] Ideker T, Galitski T, Hood L. A new approach to decoding life: systems biology. Annu Rev Genomics Hum Genet 2001;2:343–72.
[3] Chuang HY, Hofree M, Ideker T. A decade of systems biology. Annu Rev Cell Dev Biol 2010;26:721–44.
[4] Bruggeman FJ, Westerhoff HV. The nature of systems biology. Trends Microbiol 2007;15:45–50.
[5] Kimball AB, Grant RA, Wang F, Osborne R, Tiesman JP. Beyond the blot: cutting edge tools for genomics, proteomics and metabolomics analyses and previous successes. Br J Dermatol 2012;166:1–8.
[6] Zhao S, Fung-Leung WP, Bittner A, Ngo K, Liu X. Comparison of RNA-seq and microarray in transcriptome profiling of activated T cells. PLoS One 2014;9:e78644.
[7] Beck M, Schmidt A, Malmstroem J, Claassen M, Ori A, Szymborska A, et al. The quantitative proteome of a human cell line. Mol Syst Biol 2014;7.549–549.
[8] Rabilloud T. The whereabouts of 2D gels in quantitative proteomics. Methods Mol Biol 2012;893:25–35.
[9] Durmuş S, Çakır T, Özgür A, Guthke R. A review on computational systems biology of pathogen–host interactions. Front Microbiol 2015;6:235.
[10] Schleicher J, Conrad T, Gustafsson M, Cedersund G, Guthke R, Linde J. Facing the challenges of multiscale modelling of bacterial and fungal pathogen–host interactions. Brief Funct Genomics 2016:1–13.
[11] Horn F, Heinekamp T, Kniemeyer O, Pollmächer J, Valiante V, Brakhage AA. Systems biology of fungal infection. Front Microbiol 2012;3:108.
[12] Maji A, Misra R, Kumar Mondal A, Kumar D, Bajaj D, Singhal A, et al. Expression profiling of lymph nodes in tuberculosis patients reveal inflammatory milieu at site of infection. Sci Rep 2015;5:15214.
[13] Blackham S, Baillie A, Al-Hababi F, Remlinger K, You S, Hamatake R, et al. Gene expression profiling indicates the roles of host oxidative stress, apoptosis, lipid metabolism, and intracellular transport genes in the replication of hepatitis C virus. J Virol 2010;84:5404–14.
[14] Zaugg C, Monod M, Weber J, Harshman K, Pradervand S, Thomas J, et al. Gene expression profiling in the human pathogenic dermatophyte *Trichophyton rubrum* during growth on proteins. Eukaryot Cell 2009;8:241–50.
[15] Westermann AJ, Gorski SA, Vogel J. Dual RNA-seq of pathogen and host. Nat Rev Microbiol 2012;10:618–30.
[16] Schulze S, Henkel SG, Driesch D, Guthke R, Linde J. Computational prediction of molecular pathogen–host interactions based on dual transcriptome data. Front Microbiol 2015;6.
[17] Moy P, Qutob D, Chapman BP, Atkinson I, Gijzen M. Patterns of gene expression upon infection of soybean plants by *Phytophthora sojae*. Mol Plant Microbe Interact 2004;17:1051–62.
[18] Ithal N, Recknor J, Nettleton D, Hearne L, Maier T, Baum TJ, et al. Parallel genome-wide expression profiling of host and pathogen during soybean cyst nematode infection of soybean. Mol Plant Microbe Interact 2007;20:293–305.
[19] Tierney L, Linde J, Müller S, Brunke S, Molina JC, Hube B, et al. An interspecies regulatory network inferred from simultaneous RNA-seq of *Candida albicans* invading innate immune cells. Front Microbiol 2012;3.
[20] AbuOun M, Suthers PF, Jones GI, Carter BR, Saunders MP, Maranas CD, et al. Genome scale reconstruction of a salmonella metabolic model: comparison of similarity and differences with a commensal *Escherichia coli* strain. J Biol Chem 2009;284:29480–8.
[21] Bumann D. System-level analysis of *Salmonella* metabolism during infection. Curr Opin Microbiol 2009;12:559–67.
[22] Kim YM, Schmidt BJ, Kidwai AS, Jones MB, Deatherage Kaiser BL, Brewer HM, et al. *Salmonella* modulates metabolism during growth under conditions that induce expression of virulence genes. Mol Biosyst 2013;9:1522.
[23] Navid A, Almaas E. Genome-level transcription data of *Yersinia pestis* analyzed with a new metabolic constraint-based approach. BMC Syst Biol 2012;6:150.
[24] Mendum TA, Wu H, Kierzek AM, Stewart GR. Lipid metabolism and type VII secretion systems dominate the genome scale virulence profile of *Mycobacterium tuberculosis* in human dendritic cells. BMC Genomics 2015;16:372.
[25] Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, et al. PATRIC, the bacterial bioinformatics database and analysis resource. Nucleic Acids Res 2014;42:D581–91.
[26] Urban M, Pant R, Raghunath A, Irvine AG, Pedro H, Hammond-Kosack KE. The Pathogen–Host Interactions database (PHI-base): additions and future developments. Nucleic Acids Res 2015;43:D645–55.
[27] Xiang Z, Tian Y, He Y. PHIDIAS: a pathogen–host interaction data integration and analysis system. Genome Biol 2007;8:R150.
[28] Kumar R, Nanduri B. HPIDB—a unified resource for host–pathogen interactions. BMC Bioinformatics 2010;11:S16.
[29] Durmus Tekir S, Cakir T, Ardic E, Sayilirbas AS, Konuk G, Konuk M, et al. PHISTO: pathogen–host interaction search tool. Bioinformatics 2013;29:1357–8.
[30] Shoaie S, Karlsson F, Mardinoglu A, Nookaew I, Bordel S, Nielsen J. Understanding the interactions between bacteria in the human gut through metabolic modeling. Sci Rep 2013;3:2532.
[31] Shoaie S, Nielsen J. Elucidating the interactions between the human gut microbiota and its host through metabolic modeling. Front Genet 2014;5.
[32] Gao YD, Zhao Y, Huang J. Metabolic modeling of common *Escherichia coli* strains in human gut microbiome. Biomed Res Int 2014;2014:1–11.
[33] Sadhukhan PP, Raghunathan A. Investigating host–pathogen behavior and their interaction using genome-scale metabolic network models. Methods Mol Biol 2014;1184:523–62.
[34] Ji B, Nielsen J. From next-generation sequencing to systematic modeling of the gut microbiome. Front Genet 2015;6:219.
[35] Heinken A, Sahoo S, Fleming RMT, Thiele I. Systems-level characterization of a host–microbe metabolic symbiosis in the mammalian gut. Gut Microbes 2013;4:28–40.
[36] Shoaie S, Ghaffari P, Kovatcheva-Datchary P, Mardinoglu A, Sen P, Pujos-Guillot E, et al. Quantifying diet-induced metabolic changes of the human gut microbiome. Cell Metab 2015;22:320–31.
[37] Strimbu K, Tavel JA. What are biomarkers? Curr Opin HIV AIDS 2010;5:463–6.
[38] Rogers TR, Morton CO, Springer J, Conneally E, Heinz W, Kenny C, et al. Combined real-time PCR and galactomannan surveillance improves diagnosis of invasive aspergillosis in high risk patients with haematological malignancies. Br J Haematol 2013;161:517–24.
[39] Lee H. Procalcitonin as a biomarker of infectious diseases. Korean J Intern Med 2013;28:285.

[40] Poddar B, Gurjar M, Singh S, Singh R, Baronia A, Aggarwal A, et al. Procalcitonin kinetics as a prognostic marker in severe sepsis/septic shock. Indian J Crit Care Med 2015;19:140.
[41] Maertzdorf J, Repsilber D, Parida SK, Stanley K, Roberts T, Black G, et al. Human gene expression profiles of susceptibility and resistance in tuberculosis. Genes Immun 2011;12:15–22.
[42] Shahabi V, Berman D, Chasalow SD, Wang L, Tsuchihashi Z, Hu B, et al. Gene expression profiling of whole blood in ipilimumab-treated patients for identification of potential biomarkers of immune-related gastrointestinal adverse events. J Transl Med 2013;11:75.
[43] Tsalik EL, Henao R, Nichols M, Burke T, Ko ER, McClain MT, et al. Host gene expression classifiers diagnose acute respiratory illness etiology. Sci Transl Med 2016;8. 322ra11–322ra11.
[44] Dix A, Hünniger K, Weber M, Guthke R, Kurzai O, Linde J. Biomarker-based classification of bacterial and fungal whole-blood infections in a genome-wide expression study. Front Microbiol 2015;6:171.
[45] Reif DM, Motsinger-Reif AA, McKinney BA, Rock MT, Crowe JE, Moore JH. Integrated analysis of genetic and proteomic data identifies biomarkers associated with adverse events following smallpox vaccination. Genes Immun 2009;10:112–9.
[46] Yousef M, Ketany M, Manevitz L, Showe LC, Showe MK. Classification and biomarker identification using gene network modules and support vector machines. BMC Bioinformatics 2009;10:337.
[47] Breiman L. Random forests. Mach Learn 2001;45:5–32.
[48] Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995;20:273–97.
[49] Altman NS. An introduction to kernel and nearest-neighbor nonparametric regression. Am Stat 1992;46:175–85.
[50] Fisher RA. The use of multiple measurements in taxonomic problems. Ann Eugen 1936;7:179–88.
[51] Lunetta KL, Hayward LB, Segal J, Van Eerdewegh P. Screening large-scale association study data: exploiting interactions using random forests. BMC Genet 2004;5:32.
[52] Lu C, Wu J, Wang H, Wang S, Diao N, Wang F, et al. Novel biomarkers distinguishing active tuberculosis from latent infection identified by gene expression profile of peripheral blood mononuclear cells. PLoS One 2011;6:e24290.
[53] Ling H, Fabbri M, Calin GA. MicroRNAs and other non-coding RNAs as targets for anticancer drug development. Nat Rev Drug Discov 2013;12:847–65.
[54] Qi Y, Cui L, Ge Y, Shi Z, Zhao K, Guo X, et al. Altered serum microRNAs as biomarkers for the early diagnosis of pulmonary tuberculosis infection. BMC Infect Dis 2012;12:384.
[55] Li LM, Hu ZB, Zhou ZX, Chen X, Liu FY, Zhang JF, et al. Serum microRNA profiles serve as novel biomarkers for HBV infection and diagnosis of HBV-positive hepatocarcinoma. Cancer Res 2010;70:9798–807.
[56] Ma Y, Vilanova D, Atalar K, Delfour O, Edgeworth J, Ostermann M, et al. Genome-wide sequencing of cellular microRNAs identifies a combinatorial expression signature diagnostic of sepsis. PLoS One 2013;8:e75918.
[57] Achkar JM, Cortes L, Croteau P, Yanofsky C, Mentinova M, Rajotte I, et al. Host protein biomarkers identify active tuberculosis in HIV uninfected and co-infected individuals. EBioMedicine 2015;2:1160–8.
[58] Zuniga J, Buendia-Roldan I, Zhao Y, Jimenez L, Torres D, Romo J, et al. Genetic variants associated with severe pneumonia in A/H1N1 influenza infection. Eur Respir J 2012;39:604–10.
[59] Abu-Maziad A, Schaa K, Bell EF, Dagle JM, Cooper M, Marazita ML, et al. Role of polymorphic variants as genetic modulators of infection in neonatal sepsis. Pediatr Res 2010;68:323–9.
[60] Perkel J. Erratum: SNP genotyping: six technologies that keyed a revolution. Nat Methods 2008;5. 575–575.
[61] Werner T. Next generation sequencing in functional genomics. Brief Bioinform 2010;11:499–511.
[62] Hecker M, Lambeck S, Toepfer S, van Someren E, Guthke R. Gene regulatory network inference: data integration in dynamic models—a review. Biosystems 2009;96:86–103.
[63] Stuart JM. A gene-coexpression network for global discovery of conserved genetic modules. Science 2003;302:249–55.
[64] Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Favera R, et al. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics 2006;7:S7.
[65] Meyer PE, Kontos K, Lafitte F, Bontempi G. Information-theoretic inference of large transcriptional regulatory networks. EURASIP J Bioinforma Syst Biol 2007;2007:1–9.
[66] Faith JJ, Hayete B, Thaden JT, Mogno I, Wierzbowski J, Cottarel G, et al. Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. PLoS Biol 2007;5:e8.
[67] Zoppoli P, Morganella S, Ceccarelli M. TimeDelay-ARACNE: reverse engineering of gene networks from time-course data by an information theoretic approach. BMC Bioinformatics 2010;11:154.
[68] Tibshirani R. Regression shrinkage and selection via the Lasso. J R Stat Soc Ser B 1996;58:267–88.
[69] Efron B, Hastie T, Johnstone I, Tibshirani R. Least angle regression. Ann Stat 2004;32:407–99.
[70] Hecker M, Goertsches R, Engelmann R, Thiesen HJ, Guthke R. Integrative modeling of transcriptional regulation in response to antirheumatic therapy. BMC Bioinformatics 2009;10:262.
[71] Jeong H, Mason SP, Barabási AL, Oltvai ZN. Lethality and centrality in protein networks. Nature 2001;411:41–2.
[72] Potapov AP, Voss N, Sasse N, Wingender E. Topology of mammalian transcription networks. Genome Inform 2005;16:270–8.
[73] Altwasser R, Linde J, Buyko E, Hahn U, Guthke R. Genome-wide scale-free network inference for *Candida albicans*. Front Microbiol 2012;3:1–10.
[74] Mjolsness E. On cooperative quasi-equilibrium models of transcriptional regulation. J Bioinform Comput Biol 2007;5:467–90.
[75] Müssel C, Hopfensitz M, Kestler HA. BoolNet—an R package for generation, reconstruction and analysis of Boolean networks. Bioinformatics 2010;26:1378–80.
[76] Hartemink AJ, Gifford DK, Jaakkola TS, Young RA. Using graphical models and genomic expression data to statistically validate models of genetic regulatory networks. Pacific Symp Biocomput 2001:422–33.
[77] Vlaic S, Schmidt-Heck W, Matz-Soja M, Marbach E, Linde J, Meyer-Baese A, et al. The extended TILAR approach: a novel tool for dynamic modeling of the transcription factor network regulating the adaption to *in vitro* cultivation of murine hepatocytes. BMC Syst Biol 2012;6:147.
[78] Weber M, Henkel SG, Vlaic S, Guthke R, van Zoelen EJ, Driesch D. Inference of dynamical gene-regulatory networks based on time-resolved multi-stimuli multi-experiment data applying NetGenerator V2.0. BMC Syst Biol 2013;7:1.
[79] Yang X, Dent JE, Nardini C. An SSystem Parameter Estimation Method (SPEM) for biological networks. J Comput Biol 2012;19:175–87.
[80] Linde J, Hortschansky P, Fazius E, Brakhage AA, Guthke R, Haas H. Regulatory interactions for iron homeostasis in *Aspergillus fumigatus* inferred by a systems biology approach. BMC Syst Biol 2012;6:6.
[81] Ghiassian SD, Menche J, Barabási AL. A DIseAse MOdule Detection (DIAMOnD) algorithm derived from a systematic analysis of connectivity patterns of disease proteins in the human interactome. PLOS Comput Biol 2015;11:e1004120.
[82] Barrenäs F, Chavali S, Alves A, Coin L, Jarvelin MR, Jörnsten R, et al. Highly interconnected genes in disease-specific networks are enriched for disease-associated polymorphisms. Genome Biol 2012;13:R46.
[83] Sharma A, Menche J, Huang CC, Ort T, Zhou X, Kitsak M, et al. A disease module in the interactome explains disease heterogeneity, drug response and captures novel pathways and genes in asthma. Hum Mol Genet 2015;24:3005–20.
[84] Gustafsson M, Edström M, Gawel D, Nestor CE, Wang H, Zhang H, et al. Integrated genomic and prospective clinical studies show the importance of modular pleiotropy for disease susceptibility, diagnosis and treatment. Genome Med 2014;6:17.
[85] Menche J, Sharma A, Kitsak M, Ghiassian SD, Vidal M, Loscalzo J, et al. Disease networks. Uncovering disease–disease relationships through the incomplete interactome. Science 2015;347:1257601.
[86] Wu G, Stein L. A network module–based method for identifying cancer prognostic signatures. Genome Biol 2012;13:R112.
[87] Remmele CW, Luther CH, Balkenhol J, Dandekar T, Müller T, Dittrich MT. Integrated inference and evaluation of host–fungi interaction networks. Front Microbiol 2015;6:764.
[88] Smeekens SP, van de Veerdonk FL, Netea MG. An omics perspective on *Candida* infections: towards next-generation diagnosis and therapy. Front Microbiol 2016;7. 1664–302X.
[89] Yoon SH, Park YK, Kim JF. PAIDB v2.0: exploration and analysis of pathogenicity and resistance islands. Nucleic Acids Res 2015;43:D624–30.
[90] Zhang H, Gustafsson M, Nestor C, Chung KF, Benson M. Targeted omics and systems medicine: personalising care. Lancet Respir Med 2014;2:785–7.
[91] Caie PD, Schuur K, Oniscu A, Mullen P, Reynolds PA, Harrison DJ. Human tissue in systems medicine. FEBS J 2013;280:5949–56.
[92] Auffray C, Chen Z, Hood L. Systems medicine: the future of medical genomics and healthcare. Genome Med 2009;1:2.
[93] Reyes-Palomares A, Rodríguez-López R, Ranea JAG, Jiménez FS, Medina MA. Global analysis of the human pathophenotypic similarity gene network merges disease module components. PLoS One 2013;8:e56653.
[94] Fischer K, Kettunen J, Würtz P, Haller T, Havulinna AS, Kangas AJ, et al. Biomarker profiling by nuclear magnetic resonance spectroscopy for the prediction of all-cause mortality: an observational study of 17,345 persons. PLoS Med 2014;11:e1001606.
[95] Akinkuolie AO, Buring JE, Ridker PM, Mora S. A novel protein glycan biomarker and future cardiovascular disease events. J Am Heart Assoc 2014;3. e001221–e001221.
[96] Ritchie SC, Würtz P, Nath AP, Abraham G, Havulinna AS, Fearnley LG, et al. The biomarker GlycA is associated with chronic inflammation and predicts long-term risk of severe infection. Cell Syst 2015;1:293–301.
[97] Capobianco E. Ten challenges for systems medicine. Front Genet 2012;3:193.
[98] Gustafsson M, Nestor CE, Zhang H, Barabási AL, Baranzini S, Brunak S, et al. Modules, networks and systems medicine for understanding disease and aiding diagnosis. Genome Med 2014;6:82.
[99] Pearson E. Personalized medicine in diabetes: the role of 'omics' and biomarkers. Diabet Med 2016. In press.
[100] Hu ZZ, Huang H, Wu CH, Jung M, Dritschilo A, Riegel AT, et al. Omics-based molecular target and biomarker identification. Methods Mol Biol 2011;719:547–71.
[101] Mardis ER. Anticipating the 1,000 dollar genome. Genome Biol 2006;7:112.
[102] Guest PC, Gottschalk MG, Bahn S. Proteomics: improving biomarker translation to modern medicine? Genome Med 2013;5:17.

[103] Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, et al. Global quantification of mammalian gene expression control. Nature 2011;473:337–42.
[104] Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat Rev Genet 2012;13:227–32.
[105] Cunha C, Aversa F, Lacerda JF, Busca A, Kurzai O, Grube M, et al. Genetic PTX3 deficiency and aspergillosis in stem-cell transplantation. N Engl J Med 2014;370:421–32.
[106] Lu Y, Goldstein DB, Angrist M, Cavalleri G. Personalized medicine and human genetic diversity. Cold Spring Harb Perspect Med 2014;4. a008581–a008581.
[107] Liu LY, Yang T, Ji J, Wen Q, Morgan AA, Jin B, et al. Integrating multiple 'omics' analyses identifies serological protein biomarkers for preeclampsia. BMC Med 2013;11:236.
[108] Vodovotz Y, An G. Translational systems biology. Amsterdam: Elsevier; 2015.
[109] Vodovotz Y, Constantine G, Faeder J, Mi Q, Rubin J, Bartels J, et al. Translational systems approaches to the biology of inflammation and healing. Immunopharmacol Immunotoxicol 2010;32:181–95.
[110] Vodovotz Y, Csete M, Bartels J, Chang S, An G. Translational systems biology of inflammation. PLoS Comput Biol 2008;4:e1000014.
[111] Cedersund G, Roll J. Systems biology: model based evaluation and comparison of potential explanations for given biological data. FEBS J 2009;276:903–22.
[112] Pienaar E, Dartois V, Linderman JJ, Kirschner DE. *In silico* evaluation and exploration of antibiotic tuberculosis treatment regimens. BMC Syst Biol 2015;9:79.
[113] Nyman E, Rajan MR, Fagerholm S, Brannmark C, Cedersund G, Stralfors P. A single mechanism can explain network-wide insulin resistance in adipocytes from obese patients with type 2 diabetes. J Biol Chem 2014;289:33215–30.
[114] Vogt H, Hofmann B, Getz L. The new holism: P4 systems medicine and the medicalization of health and life itself. Med Heal Care Philos 2016:1–17.
[115] Saqi M, Pellet J, Roznovat I, Mazein A, Ballereau S, De Meulder B, et al. Systems medicine: the future of medical genomics, healthcare, and wellness. Methods Mol Biol 2016;1386:43–60.